Computers and Mathematics with Applications 62 (2011) 4514–4524

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches

Mu-Yen Chen

Department of Information Management, National Taichung Institute of Technology, Taichung 404, Taiwan, ROC

ARTICLE INFO

Article history:
Received 7 September 2010
Received in revised form 25 June 2011
Accepted 10 October 2011

Keywords:
Decision tree classification
Support vector machine
Financial bankruptcy prediction

ABSTRACT

In this paper, we compare some traditional statistical methods for predicting financial distress to some more "unconventional" methods, such as decision tree classification, neural networks, and evolutionary computation techniques, using data collected from 200 Taiwan Stock Exchange Corporation (TSEC) listed companies. Empirical experiments were conducted using a total of 42 ratios including 33 financial, 8 non-financial and 1 combined macroeconomic index, using principle component analysis (PCA) to extract suitable variables.

This paper makes four critical contributions: (1) with nearly 80% fewer financial ratios by the PCA method, the prediction performance is still able to provide highly-accurate forecasts of financial bankruptcy; (2) we show that traditional statistical methods are better able to handle large datasets without sacrificing prediction performance, while intelligent techniques achieve better performance with smaller datasets and would be adversely affected by huge datasets; (3) empirical results show that C5.0 and CART provide the best prediction performance for imminent bankruptcies; and (4) Support Vector Machines (SVMs) with evolutionary computation provide a good balance of high-accuracy short- and long-term performance predictions for healthy and distressed firms. Therefore, the experimental results show that the Particle Swarm Optimization (PSO) integrated with SVM (PSO–SVM) approach could be considered for predicting potential financial distress.

1. Introduction

Methods for predicting bankruptcy of financial firms became an important issue in the 1960s and have been widely investigated since [1]. Increased emphasis on this topic could be taken as an indicator of the degree of development and robustness of a given country's economy [2]. The high individual, economic, and social costs inherent in corporate failures or bankruptcies have prompted efforts to provide better insight into and prediction of bankruptcy events [3]. Given the radical change of globalization, more accurate forecasting of corporate financial distress would provide useful information for decision-makers, such as stockholders, creditors, governmental officials, and even the general public. In fact, corporate bankruptcies can be caused by many factors such as wrong investment decisions, a poor investment environment, low cash flow and so on [1,4,5]. Therefore, the many current methods for predicting corporate failure must be continuously improved.

Bankruptcy prediction models can be classified into two broad categories: statistical and artificial intelligent (AI) techniques. Beaver [6] pioneered the statistical methods, followed by Altman [1] who applied multi-discriminant analysis (MDA), and also developed stochastic models such as logit [5] and probit [7]. However the practical application of statistical

E-mail address: mychen@ntit.edu.tw.

0898-1221/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2011.10.030

methods is limited by their inherent strict assumptions such as linearity, normality, independence among predictor variables and pre-existing functional forms relating to the criterion variable and the predictor variable [8]. Over the past decade, a number of studies have applied artificial intelligent techniques to bankruptcy prediction. Currently, these techniques include (i) decision trees including Interactive Dichotomiser 3 (ID3) [9], C5.0 [10] and classification and regression tree (CART) [11]; (ii) different artificial neural network (ANN) architectures including multi-layer perception (MLP) [12], self-organizing map (SOM) [13] and learning vector quantization (LVQ) [14], (iii) evolutionary approaches including genetic algorithms (GA) [15] and a newer evolutionary technique – particle swarm optimization (PSO); and (iv) other intelligent techniques including Support Vector Machines (SVMs) [8].

Among the intelligent techniques, decision trees form a part of 'machine learning' which is an important area of artificial intelligence [16]. Most decision tree algorithms are used for solving classification problems. Decision trees could be used as a partitioning method to induce the rules of a given dataset and act as a prediction model for future datasets applying the recursive partitioning algorithm (RPA) for predicting bankruptcy in firms proposed by Marais et al. [17]. Frydman et al. [18] applied RPA to bankruptcy prediction and compared it with the MDA. Cho et al. [19] also compared decision trees and case-based reasoning for bankruptcy prediction. C5.0 is a new decision tree algorithm developed based on C4.5 by Quinlan [20]. It includes all functionality of C4.5 and applies the boosting technology for improved accuracy in sample identification. In the ANN approach, MLP has extensive applications in financial services. Since SOM and LVQ are not often used in the financial domain, we investigate the bankruptcy prediction performance of both models. In the evolutionary approach, GA and PSO can enhance the capability and probability of finding global optima and optimizing parameters in the results. Recently, SVM has been used in financial prediction applications such as credit ratings, time series predictions and the detection of insurance claim fraud [21]. Therefore, the purpose of this paper is to compare these various classification techniques to predict financial distress.

In this research, the comparison models use the linear discriminant analysis (LDA), logistic regression (LR), C5.0, CART, SOM, LVQ, SOM, GA, and PSO techniques. The main objectives of this paper are to (1) construct financial distress prediction models from classification techniques, (2) increase the accuracy of these models using financial, non-financial, and macroeconomic ratios, (3) compare the accuracy of traditional statistical, intelligent, and evolutionary computation approaches, and (4) expand these models to form a financial distress prediction system to provide information to investors and monitoring organizations. The data employed in our study was collected from Taiwan Stock Exchange Corporation (TSEC) databases.

The rest of the paper is organized as follows: Section 2 reviews the literature on the comparative analysis of statistical, intelligent, and evolutionary computation approaches. Section 3 provides a brief description of the data organization and the research model. Section 4 presents research results and analysis. Section 5 presents our conclusions.

2. Literature review

We are now going to provide a glimpse into the literature concerning the main statistical and soft computing techniques that so far have been used to analyze distress situations and the bankruptcy prediction problem. In particular, we will focus on LDA, LR, C5.0, CART, SOM, LVQ, SVM, GA–SVM, PSO–SVM techniques and brief details of each of the techniques in this section.

2.1. Statistical techniques

Discriminant analysis (DA) is commonly used to classify a set of observations into predefined classes. Cambell [22] and Fung [23] suggested a diagnostic method in LDA for detecting the possible influential observations based on the influence function. LDA is a classification method which assumes that data in each class are Gaussian-distributed and that there is a unique covariance matrix for each class. LDA also maps the data in a transformed space formed by the eigenvectors of the pooled covariance matrix. Hence a new instance can be classified simply by mapping it in the transformed space and assigning the class of the closest centroid, a process known as linear decision boundaries. However, LDA has some pitfalls. First, linear decision boundaries are inadequate to deal with the Small Sample Size (S3) problem which occurs when the total number of training samples is smaller than the dimensionality of the feature vector. In this situation, the within-class scatter matrix becomes singular; it becomes impossible for LDA to handle the linear decision boundaries. Second, using a single class prototype may prove insufficient and, several prototypes are more appropriate in many situations. Third, we may have too many correlated predictors. Given these issues, we have replaced discriminant analysis with LR, which is much more flexible in its assumptions.

LR is a regression method for predicting a dichotomous dependent variable. Unlike LDA, LR does not require that independent variables be normally distributed or linearly related, nor does it require equal variance within each group [24]. In LR models, the dependent variable is always in categorical form and has two or more levels. Independent variables may be in numerical or categorical form. Recently, many researchers have applied LR to predict financial bankruptcies. Laitinen and Laitinen [25] used Taylor's model in bankruptcy prediction and evaluated the application of the LR model to data from the Compustat database. Premachandra et al. [26] found that LE outperformed data envelopment analysis (DEA) in predicting corporate bankruptcies. LDA and LR are useful for benchmarking other techniques, and are included in our experiment.